STATISTICAL EXTRAPOLATION

Dennis Randolph Oldson

United States
Naval Postgraduate School

THESIS

STATISTICAL EXTRAPOLATION

by

Dennis Randolph Oldson

Thesis Advisor: Sydney R. Parker

September 1971

Approved for public release; distribution unlimited.

Statistical Extrapolation

by

Dennis Randolph Oldson
Lieutenant, United States Navy
B.S.E.E., University of Colorado, 1965

Submitted in partial fulfillment of the
requirements for the degree of

ELECTRICAL ENGINEER

from the

NAVAL POSTGRADUATE SCHOOL
September 1971

ABSTRACT

The mathematics of extrapolating known statistics of components to the probability density function of a system's performance measure is considered. Quadrature sum integration schemes for evaluating the resulting required integration are examined, and alternate integral approximation schemes are developed utilizing Monte Carlo methods. A simple electrical circuit example illustrates the use of these techniques.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The desirability of being able to statistically predict a system's performance measure is obvious when one considers the mass production of systems. Since performance measures of certain values may not be desirable, it is generally unprofitable to produce such systems when the probability that these undesirable values can exist is beyond the limits determined by the requirements of the situation.

Presently performance data for systems are obtained in several ways: (1) A worst case analysis [1] to determine if undesirable performance will occur within the parameter tolerance; (2) A moment method [1] in which the performance measure is sampled so that its mean, variance, and higher order moments are determined for approximation by known distribution functions; (3) A functional formulas method [2] in which performance measure's distributions are obtained from breaking performance measures into a series of products and sums for individual integration; (4) Monte Carlo methods [3, 10] in which performance measure distributions are generated by creation of random processes by which the parameters of the random process lead to an approximation to the desired distribution.

The purpose of this thesis is to investigate the mathematics required in obtaining a performance measure and attempt to arrive at a reasonably fast and accurate method of

extrapolating the statistics of a system's component variables to create the distribution of the desired performance measure.

Chapter II examines the mathematical form and statistical nature of the performance measure for implicit and explicit functions and presents two digital computer integration schemes using Gauss quadrature sums for evaluating the resulting required integration.

Chapter III investigates crude Monte Carlo methods for determining performance measures, and the confidence limits involved in the method. Then Monte Carlo schemes to estimate single integrals are examined. The multiple integral extensions of the single integral estimates are then derived.

Chapter IV discusses the relative merits of the methods outlined in Chapters II and III, and gives a comparison of the results of these methods applied to a simple example consisting of an electrical circuit in which the component values are the random variables.

II. STATISTICAL ANALYSIS

The probability density function of a system's performance measure can be determined mathematically from knowledge of the functional relation between the performance measure and the system component-variables, and knowledge of the joint probability density function of the component-variables. Most often this requires integrating a complex function over several variables. Although the integration can be performed analytically in some cases, digital computer integration schemes must normally be utilized.

This chapter examines the mathematics of determining the probability density function of a performance measure, first for performance measures which are explicit functions of the component variables, then for implicit performance measures. Two integration schemes are then presented for performing the integration on a digital computer.

A. EXPLICIT CASE

Consider a performance measure, Z as a function of the system's component-variables, $\underline{X}^T$, where $\underline{X}^T$ is the transpose of an n-vector of elements $X_1, X_2, \ldots, X_n$. It is assumed that the X_i's are statistically independent, so that the joint probability density function of $\underline{X}^T$ is given by the product of the probability density functions of the individual X_i's.

The probability density function of $Z = g(\underline{X}^T)$ can be written as [7]

$$p_Z(z) = \int_V p_{\underline{X}^T,Z}(\underline{x}^T,z)\,dV \tag{1}$$

where $p_{\underline{X}^T,Z}(\underline{x}^T,z)$ is the joint probability density function for $\underline{X}^T$ and Z, V is the n-dimensional space determined by the component-variables, g is the functional relation between Z and the component variables, $\underline{X}^T$, and $dV = dx_1 dx_2 \ldots dx_n$.

If a transformation of variables is defined such that

$$\underline{Y}^T = (X_2, X_3, \ldots, X_n) \;;\; \underline{X}^T = (X_1 \underline{Y}^T) \tag{2}$$

$$X_1 = f(Z, \underline{Y}^T) \tag{3}$$

where f is the functional relation between the component-variable X_1 and the performance measure Z, equation (1) becomes [7, 9]

$$p_Z(z) = \int_U p_{\underline{X}^T}[f(z,\underline{y}^T),\underline{y}^T]\,|J(z)|\,dU \tag{4}$$

where U is the n-1 dimensional space determined by the n-1 component-variables $\underline{Y}^T$, $dU = dy_1 dy_2 \ldots dy_{n-1}$, $y_j = x_{j+1}$, and J(z) is the Jacobian of the transformation (2) and (3).

J(z) can be written as

$$J(z) = \begin{vmatrix} \frac{\partial x_1}{\partial z} & \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_{n-1}} \\ \frac{\partial x_2}{\partial z} & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \frac{\partial x_n}{\partial z} & \cdot & \cdots & \frac{\partial x_n}{\partial y_{n-1}} \end{vmatrix}$$

$$= \begin{vmatrix} \frac{\partial x_1}{\partial z} & 0 \quad \cdots \quad 0 \\ \cdot & \\ \cdot & \mathbf{I} \\ \cdot & \\ \frac{\partial x_n}{\partial z} & \end{vmatrix} = \frac{\partial x_1}{\partial z} \tag{5}$$

where I is the n-1 by n-1 identity matrix.

Since the joint probability density function of statistically independent random variables is the product of the individual probability density functions, equation (4) becomes

$$p_Z(z) = \int_U p_{X_1}[f(z,\underline{y}^T),\underline{y}^T] \, |\partial x_1/\partial z| \, p_{\underline{Y}^T}(\underline{y}^T)\,dU \tag{6}$$

In the case where the performance measure is a known or explicit function of the component-variables, the inverse functional relation f relating X_1 to Z can usually be expressed analytically as can the partial derivative of Z with respect to X_1 or any of the other component-variables. Circumstances may require reordering the component-variables to obtain these relations in a convenient form.

As an example, consider the series R-L-C circuit in Figure 1. The current I through the circuit is

$$I(j\omega) = \frac{E(j\omega)}{R + j(\omega L - 1/\omega C)} \tag{7}$$

and is a maximum for $\omega L = 1/\omega C$. The angular frequency at which the current falls to 3 db of the maximum occurs when

$$R = \omega L - 1/\omega C \tag{8}$$

or,

$$\omega_{3db} = R/2L \pm \sqrt{R^2/4L^2 + 1/LC} \tag{9}$$

Let the performance measure Z be the upper 3 db frequency and $\underline{X}^T = (R,L,C)$, so that

$$Z = g(\underline{X}^T) = X_1/2X_2 + \sqrt{X_1{}^2/4X_2{}^2 + 1/X_2X_3} \tag{10}$$

and from (8),

$$X_1 = f(Z,\underline{Y}^T) = ZX_2 - 1/ZX_3 \tag{11}$$

so that the partial derivative is

$$\partial X_1/\partial Z = X_2 + 1/Z^2X_3 \tag{12}$$

If each of the components has a Gaussian distribution about its mean such that

$$p_{X_j} = \frac{\exp[-(x_j-\bar{x}_j)^2/2\sigma_j{}^2]}{(2\pi)^{1/2}\sigma_j},\ j=1,2,3 \tag{13}$$

where $\bar{x}_j$ is the expected or average value of the variable X_j, and σ_j is the standard deviation of the variable X_j.

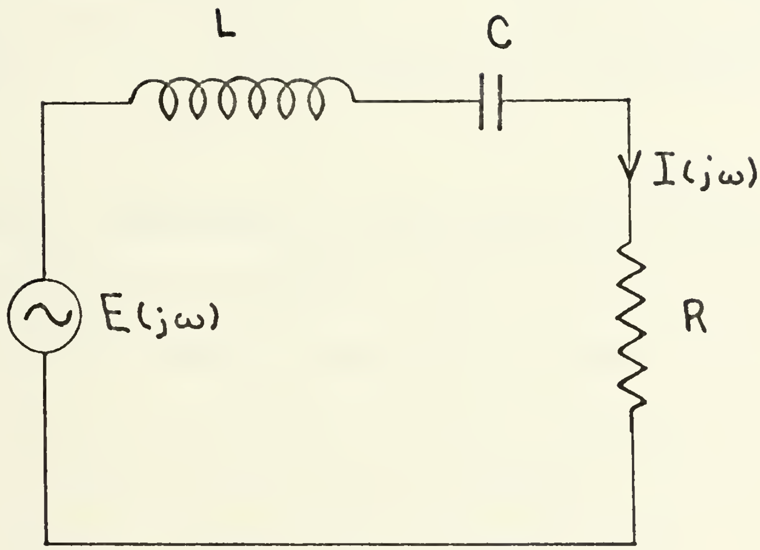

Figure 1. Series R-L-C Circuit

Substituting for X_1 in (13) for $j = 1$, $p_{X_1}[f(z,\underline{y}^T),\underline{y}^T]$ becomes

$$p_{X_1} = \frac{\exp[-(zx_2-1/zx_3-\bar{x}_1)^2/2\sigma_1^2]}{(2\pi)^{1/2}\sigma_1} \tag{14}$$

Since X_2 and X_3 are independent, $p_{\underline{y}^T}(\underline{y}^T)$ is the product of $p_{X_2}(x_2)$ and $p_{X_3}(x_3)$, and (6) becomes

$$p_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp\left[\dfrac{-(zx_2-1/zx_3-\bar{x}_1)^2}{2\sigma_1^2} - \dfrac{(x_2-\bar{x}_2)^2}{2\sigma_2^2} - \dfrac{(x_3-\bar{x}_3)^2}{2\sigma_3^2}\right]}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3[1/(x_2+1/z^2 x_3)]}\,dx_2\,dx_3 \tag{15}$$

With a-priori knowledge of the probability density functions of the component-variables, (6) or (15) can be integrated on a digital computer using one of many algorithms available. The two quadrature formulas presented below offer routines which are fast, and for the number of points at which the integrand is evaluated, two of the most accurate routines available [5, 8].

B. IMPLICIT CASE

If the performance measure is not known explicitly, but is obtained implicitly from a solution of a mathematical model of the system, relation (3) cannot be determined. One may, however, resort to the cumulative distribution obtained from (1) by [7]

$$F(z) = P[Z \leq z] = \int_{-\infty}^{z} \int_{V} p_{\underline{X}^T,Z}(\underline{x}^T, z)\, dV dz \tag{16}$$

Since

$$p_{\underline{X}^T,Z}(\underline{x}^T, z) = p_{Z/\underline{X}^T}(z/\underline{x}^T)\, p_{\underline{X}^T}(\underline{x}^T) \tag{17}$$

where $p_{Z/X^T}(z/\underline{x}^T)$ is the conditional probability of Z given that $\underline{X}^T$ has occurred; and for a given $\underline{x}^T$, $p_{Z/\underline{X}^T}(z/\underline{x}^T)$ is either 0 or 1, the cumulative distribution can be approximated by

$$F(z) = \sum_{Z \leq z} \Delta z \int_{V} p_{\underline{X}^T}(\underline{x}^T)\, dV \tag{18}$$

where the integration over the space V may be performed utilizing the quadrature formulas below with the value of $p_{\underline{X}^T}(\underline{x}^T)$ as calculated for $Z(\underline{x}^T) \leq z$ and $p_{\underline{X}^T}(\underline{x}^T) = 0$ if $Z(\underline{x}^T) > z$.

The cumulative distribution function for the performance measure can then be fitted to a polynomial, and if desired, the polynomial can be differentiated analytically to obtain the probability density function.

C. QUADRATURE SUMS FOR EVALUATING INTEGRALS

The form of the performance measure has been reduced to an integral or multiple integral which, in general, cannot be evaluated in closed form. All algorithms which approximate the value of an integral by a linear combination of values of the integrand are exact [5] for the integrand being of a

certain degree or less; and in most cases the degree is one less then the number of points at which the integrand is evaluated. Quadrature sum algorithms utilize orthoganal polynomials to arrive at the form of the linear combination of values of the integrand; and as a result are exact for the integrand being a polynomial of degree one less than twice the number of points at which the integrand is evaluated. [5]

Consider the integral

$$I = \int_a^b h(x)g(x)\,dx \tag{19}$$

where a and b are real numbers, finite or infinite, g(x) is an arbitrary function and h(x) is a particular weighting function to be described. Then by selecting a polynomial $Q_n(x)$ of degree n such that $h(x)Q_n(x)$ is orthoganal to x^m, for all m = 0, 1, ..., n-1, so that

$$\int_a^b x^m h(x)Q_n(x)\,dx = 0 \tag{20}$$

If (20) is approximated by a linear combination of the integrand

$$\int_a^b x^m h(x)Q_n(x)\,dx = \sum_{k=1}^{n} A_k {x_k}^m Q_n(x_k) \tag{21}$$

then the right hand side of (21) will be equal to zero for any set of values of A_k if the x_k's are chosen such that they are the zeros of $Q_n(x)$.

By forming the n sums

$$\int_a^b x^m h(x)\,dx = \sum_{k=1}^{n} A_k x_k^{\,m} \tag{22}$$

for $m = 0,1,\ldots,n-1$, and for the n values of x_k being the n zeros of $Q_n(x)$, the n values of A_k can be found such that

$$\int_a^b h(x)g(x)\,dx = \sum_{k=1}^{n} A_k g(x_k) \tag{23}$$

is exact for all polynomials of degree 2n-1. [5]

It can be shown that if the polynomial $Q_n(x)$ is known then the values of A_k can be found to be [5, 8].

$$A_k = 2/(1-x_k^{\,2})[dQ_n(x_k)/dx]^2 \tag{24}$$

1. Gauss Legendre Quadrature

The first obvious choice of w(x) is 1, and if the interval [a, b] can be normalized to the interval [-1, 1] by a change of variables, (23) becomes

$$\int_{-1}^{1} g(x)\,dx = \sum_{k=1}^{n} A_k g(x_k) \tag{25}$$

and the polynomial $Q_n(x)$ is a Legendre polynomial of degree n [5].

$$Q_n(x) = \frac{d^n(x^2-1)^n}{dx^n 2^n n!} \tag{26}$$

The n values of x_k can be found from the zeros of $Q_n(x)$ and the values A_k from (24) or the values can be obtained from Table II for values of n from 2 to 6 and from [5] for values of n from 2 to 48. The magnitude of the error bound can be shown to be [5],

$$e_{GL} \leq \frac{2^{2n+1}(n!)^2}{(2n+1)(2n)!} \max_x \left|\frac{df^{2n}}{dx^{2n}}\right| \tag{27}$$

2. Gauss Hermite Quadrature

As is often the case when dealing with probability density functions, a variable is Gaussian in distribution, and as a result an integral of the form

$$\int_{-\infty}^{\infty} h(x)\exp(-x^2)\,dx = \sum_{k=1}^{n} A_k h(x_k) \tag{28}$$

must be evaluated. This is the form of (19) where $w(x) = \exp(-x^2)$.

The class of Chebychev-Hermite polynomials are orthogonal with respect to $\exp(-x^2)$ over the real line, and are defined by [5]

$$Q_n(x) = (-1)^n \exp(x^2)\frac{d^n}{dx^n}[\exp(-x^2)] \tag{29}$$

The values of x_k can be found from the zero's of (29) and those of A_k from (24). However these values have been tabulated and are included in Table I for values of n from 2 to 6 and in [5] for values of n from 2 to 32.

The magnitude of the error bound for approximating (28) by a Gauss Hermite Quadrature sum can be shown to be [5]

$$e = \frac{n!(\pi)^{1/2}}{2^n(2n)!} \max_x \left|\frac{d^{2n}h(x)}{dx^{2n}}\right| \tag{30}$$

3. Application to the Example

The integrand for the example, equation (15) is of the Gauss Hermite form since the variables all have Gaussian

TABLE I

GAUSS HERMITE QUADRATURE
COEFFICIENTS AND ZEROS

Number of Evaluation Points, n	Zeros x_j	Coefficients A_j
2	$\pm$0.707 1068	0.886 2269
3	$\pm$1.224 745	0.295 4090
	0.0	1.181 635
4	$\pm$1.650 680	0.081 31284
	$\pm$0.524 6476	0.804 9141
5	$\pm$2.020 183	0.019 95324
	$\pm$0.958 5725	0.393 6193
6	$\pm$2.350 605	0.004 530010
	$\pm$1.335 849	0.157 0673
	$\pm$0.436 0774	0.724 6296

TABLE II

GAUSS LEGENDRE QUADRATURE COEFFICIENTS AND ZEROS

Number of Evaluation Points, n	Zeros x_j	Coefficients A_j
2	$\pm$0.577 3503	1.000 0000
3	$\pm$0.774 5967	0.555 5556
	0.0	0.888 8889
4	$\pm$0.861 1363	0.347 8548
	$\pm$0.339 9810	0.652 1452
5	$\pm$0.906 1798	0.236 9269
	$\pm$0.538 4693	0.478 6287
	0.0	0.568 8889
6	$\pm$0.932 4695	0.171 3245
	$\pm$0.661 2094	0.360 7616
	$\pm$0.238 6192	0.467 9139

distributions. By changing variables, let

$$x_2' = \frac{x_2-\bar{x}_2}{\sqrt{2}\sigma_2} \tag{31}$$

$$x_3' = \frac{x_3-\bar{x}_3}{\sqrt{2}\sigma_3} \tag{32}$$

$$dx_2' = \frac{dx_2}{\sqrt{2}\sigma_2} \tag{33}$$

$$dx_3' = \frac{dx_3}{\sqrt{2}\sigma_3} \tag{34}$$

Since X_2 and X_3 are independent, the double integral becomes the product of the two integrals and (15) takes the form

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(z,x_2',x_3')\exp(-x_2'^2-x_3'^2)\,dx_2'dx_3'$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} A_jA_k g(z,x_{2_j},x_{3_k}) \tag{35}$$

where

$$g(z,x_2',x_3')$$

$$= \frac{\exp[-\{z(2^{1/2}\sigma_2x_2'+\bar{x}_2)-1/z(2^{1/2}\sigma_3x_3'+\bar{x}_3)-\bar{x}_1\}^2/2\sigma_1^{\,2}]}{2^{1/2}(\pi)^{3/2}\sigma_1[2^{1/2}\sigma_2x_2'+\bar{x}_2+1/z^2(2^{1/2}\sigma_3x_3'+\bar{x}_3)]^{-1}}$$

(36)

The value of n can then be chosen as desired and the sum in (35) formed using the values of A_k and x_{2_k} and x_{3_k} from Table II or [5]. The expression (36) becomes quite formidable

to differentiate after two or more differentiations, so that one must assume that the error bound (30) is small. The approximation (35) has been performed for $n = 6$ and the mean values $R = 1000$ ohms, $L = 1.0$ millihenry, $C = 0.001$ microfarad. The standard deviation of each variable was assumed to be 3.33% of its mean value. The results are contained in Chapter IV.

Not all variables have Gaussian distributions, and most performance measures, if functions of many variables, are implicit functions. If the example performance measure is considered to have been obtained implicitly, with the same density functions for the components, the performance measure's cumulative distribution function can be determined by approximating (19) by a Gauss Legendre Quadrature sum. Since $\exp(-4.5)$ is very nearly zero, the limits of $-\infty$ and ∞ can be changed to -3σ to 3σ for each variable and by the following change of variables, normalized to the interval [-1, 1]. Let

$$x_i' = \frac{x_i - \bar{x}_i}{3\sigma_i} \quad i = 1, 2, 3 \tag{37}$$

$$dx_i' = dx_i/3\sigma_i \tag{38}$$

so that (18) becomes

$$F(z) = \sum_{Z \leq z} \int_{-1}^{1} \int_{-1}^{1} \Delta z g(x_1', x_2', x_3') dx_1' dx_2' dx_3' \tag{39}$$

where

$$g(x_1',x_2',x_3') = \frac{27\exp(-x_1'^2/2-x_2'^2/2-x_3'^2/2)}{(2\pi)^{3/2}} \tag{40}$$

The approximation of (39) by the summation

$$F(z) = \sum_{Z\leq z} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \Delta z A_i A_j A_k g(x_{1_i}', x_{2_j}', x_{3_k}') \tag{41}$$

just requires the n values of A_k and the n values of x_{i_k}' and evaluating g at these values of x_{i_k}'. It must be remembered that g takes the form of (40) only for $Z(\underline{x}^T) \leq z$ for some z; otherwise $g(\underline{x}') = 0$.

Equation (41) is evaluated for n = 6 and the same values of the components as above for the Gauss Hermite example, and the results are contained in Chapter IV for comparison. As in the Gauss Hermite example, the error is assumed to be small due to the factorials in the denominator of (27).

III. MONTE CARLO METHODS

This chapter examines the most crude of the Monte Carlo techniques for obtaining statistics on a performance measure, to methods which attempt to approximate single integrals in such a manner that the approximation is reasonably accurate. The approximations are then extended to apply to multiple integrals in such a manner that fewer evaluations of the integrand are made than in conventional methods.

Monte Carlo methods consist of solving problems of a computational nature by constructing a random process for the problem, such that the parameters of the random process are the desired quantities of the problem.

Random processes usually imply random variables which must be generated in a manner as to represent typical values from the variable's probability density function. Random number generation is then an important aspect in any Monte Carlo method of approximation. Although random number generation is a field of its own, Appendix A treats the subject suitably for purposes of this thesis.

A. CRUDE MONTE CARLO

The most crude form of Monte Carlo is that of taking samples of the performance measure to obtain an expected value and perhaps a variance of the performance measure. If the component-variables are generated in such a way so that the values are representative of their distribution functions, and

for each set of values $\underline{X}^i$ of component variables generated, the performance measure $Z_i = Z(\underline{X}^i)$ is evaluated, then the expected or average value of the performance measure is [7]

$$\bar{Z} = E(Z_i) = \frac{1}{n} \sum_{i=1}^{n} Z_i \tag{42}$$

where n is the number of samples.

While for sufficiently large n, $\bar{Z}$ is an accurate estimate of the average value of the performance measure, no knowledge of the form of the probability density function of Z is derived; and consequently no knowledge of the likelihood of other values of Z is obtained.

A more useful scheme would be to order the sample values of Z_i as follows to obtain a cumulative distribution for the performance measure.

Consider a random variable S such that if a value of the performance measure Z is sampled, and the sample value Z_i is less than an arbitrary value, say a, then $S_i = 1$; and if Z_i is greater than the value a, $S_i = 0$. Then S represents a random process. If the values of the component-variables $\underline{X}^T$ are sampled from their density function as above, then the cumulative distribution function for Z can be approximated by

$$F(a) = P[Z \leq a] = \frac{1}{n} \sum_{i=1}^{n} S_i \tag{43}$$

where n is the number of times the performance measure is sampled.

Since each of the component-variables are sampled independently for each sample value Z_i, the number of samples is not directly dependent on the number of variables.

This crude method yields a cumulative distribution function which may be fitted to a polynomial and differentiated analytically to obtain the probability density function for the performance measure Z.

As an example of the two methods, consider the example of Chapter II. Since the three variables X_1, X_2, and X_3 are assumed to be Gaussian, three variables w_1, w_2, and w_3 are generated by the method in Appendix A to approximate Gaussian distributions, each with 0 mean and variance of 1. The values of the samples x_1, x_2, x_3 are then

$$x_i = w_i\sigma_i + \bar{x}_i \ , \quad i = 1,2,3 \tag{44}$$

where σ_i and $\bar{x}_i$ are the standard deviation and expected value, respectively of the variable w_i. These values are then used to evaluate the performance measure. If the expected or average value of the performance measure is desired, the calculated values of the performance are averaged as in (42). If a cumulative distribution is desired, then for k values of a_j, j=1,2,...,k the random variable S defined above is summed to form $F(a_j)$. These values for the example are listed in Chapter IV, and a computer program listing for the computation is provided in Appendix B.

B. CONFIDENCE INTERVALS

The crude Monte Carlo method above, as with all Monte Carlo methods, yields results which are as accurate as desired within a probability or confidence determined by the desired accuracy, and the method of approximation used. What

follows is a statistical estimate of the error introduced by use of Monte Carlo approximations.

If A is an event (such as the event that the values of a performance measure is less than some specified value), and R is a random variable, such that $R_i = 1$ if A occurs on the ith sample, and $R_i = 0$ otherwise, and if T is a random variable such that

$$T = \sum_{i=1}^{n} R_i \tag{45}$$

for n samples, then T is the number of times that the event A occurs in n samples. The expected value of T is

$$E(T) = E(\sum_{i=1}^{n} R_i) = nE(R_i) = np \tag{46}$$

where p is the probability of the event A occurring. The variance of T is

$$V(T) = V(\sum_{i=1}^{n} R_i) = n^2\sigma^2 \tag{47}$$

where σ^2 is the variance of the event A.

From Chebychev's inequality [7],

$$P = P[|T-np| \leq d] \geq 1 - \sigma^2/nd^2 \tag{48}$$

where d is an arbitrarily small number. Equation (48) states that for a fixed confidence of $|T-np|$ being less than a small value d, the number of samples to achieve the accuracy, d, is bounded by

$$n \geq \frac{\sigma^2}{d^2(1-P)} \tag{49}$$

where P is the desired confidence or probability.

Another way of looking at (48) is that for a fixed confidence P, the error of the estimation varies as $\sigma n^{-1/2}$. If the variance of the random variable can be reduced to 0, the confidence of the estimation could be 1 for an error of 0.

For crude Monte Carlo, the random variable S, which takes on values of 1 whenever the sampled performance measure is less then a specified value, and values of 0 otherwise, has a binomial distribution. If p is the expected value of S, then the variance of S is $\sigma^2 = p(1-p)$ [7]. Thus for a confidence of $P = 0.9$ of being within $d=0.01$ of np for crude Monte Carlo,

$$n \geq \frac{p(1-p)}{(0.0001)(0.1)} = p(1-p)10^5 \tag{50}$$

Since the maximum value of $p(1-p)$ is 0.25, inequality (50) states that n must be greater than twenty-five thousand in order to insure the accuracy of 0.01 with a confidence of 0.9.

C. VARIANCE REDUCTION BY ESTIMATING INTEGRALS

As was seen in Chapter I, the difficulty of determining a performance measure's probability density function lies in evaluating the integral (1). Two routines were presented which approximated the integral with some accuracy. The crude Monte Carlo method illustrated above can be thought of as an approximation to an integral. Other methods exist, however which reduce the variance by hundreds of thousands over crude Monte Carlo [3]. These methods are presented below for single integrals and then a general extension is made for multiple integrals.

If I is the estimator for the integral, $\int_0^1 f(x)\,dx$, where x is a variable normalized to the interval [0, 1], then by letting

$$I = \frac{f(x)}{p_Y(y)} \tag{51}$$

where $p_Y(y)$ is the probability density function of the random variable Y and f(y) is the value of the function f at some point y sampled from $p_Y(y)$, the expected value of I is

$$E(I) = E\left[\frac{f(y)}{p_Y(y)}\right] = \int_0^1 f(y)\,dy \tag{52}$$

It is then necessary to select the probability density function $p_Y(y)$ in such a manner as to minimize the variance.

Define a function $f^*(x)$ such that the function which is $f(x)-f^*(x)$ is a straight line passing through the end-points of f(x) on the interval [0, 1], as shown in Figure 2. That is,

$$f^*(x) = f(x) - (1-x)f(0) - xf(1) \tag{53}$$

The difference function $f(x)-f^*(x)$ can be readily integrated and can be used as a first approximation to $\int_0^1 f(x)\,dx$ as

$$\int_0^1 [f(x)-f^*(x)]\,dx = 1/2f(0) + 1/2f(1) \tag{54}$$

Now if I^* is defined as

$$I^* = \frac{f^*(y)}{p_Y(y)} \tag{55}$$

it becomes necessary to sample $f^*(y)$ to get the estimate from (53), (54) and (55) for I

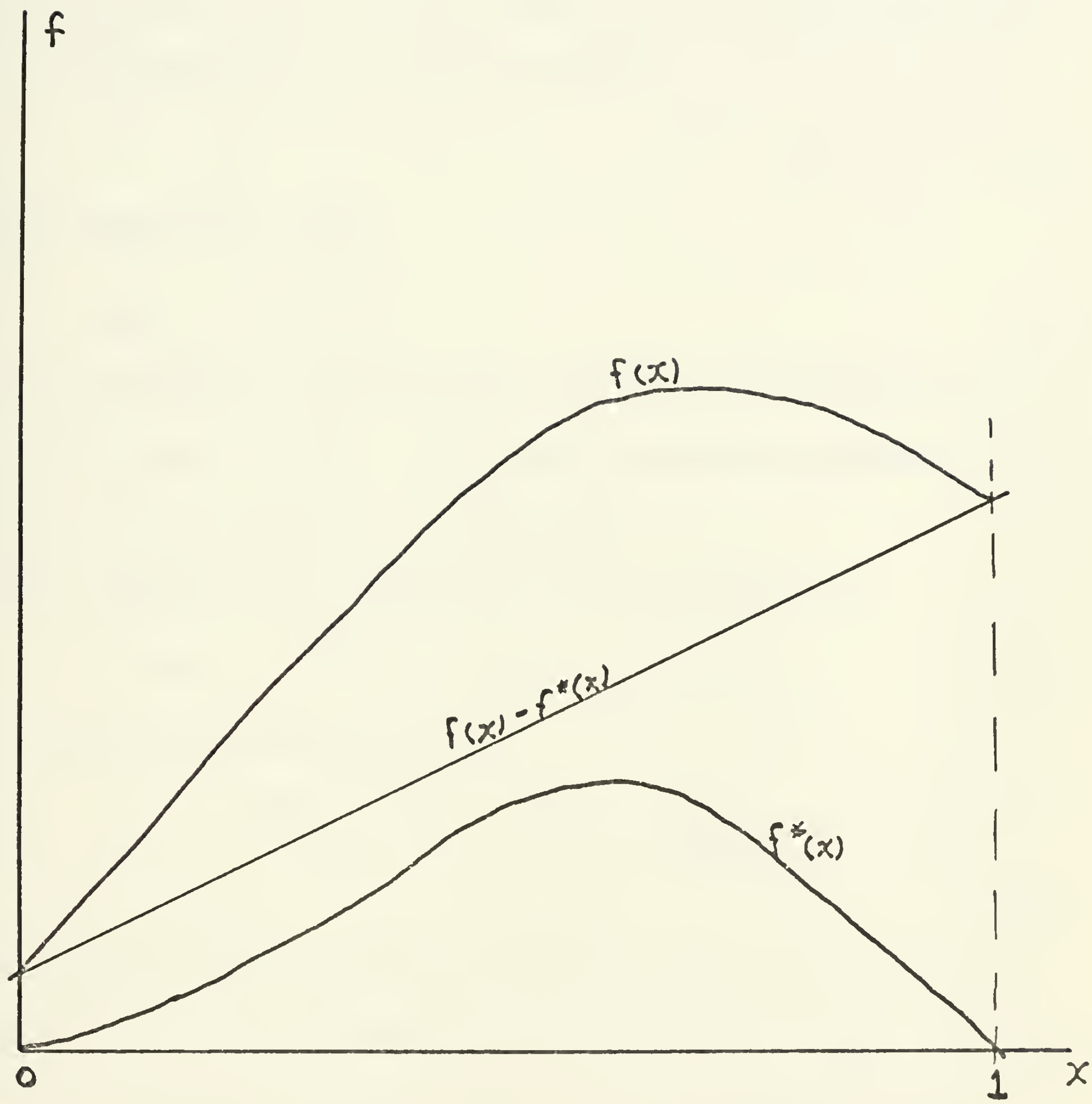

Figure 2. Difference Curve Estimation.

$$I = \frac{f(0) + f(1)}{2} + \frac{f(y) - (1-y)f(0) - y(1)}{p_Y(y)} \tag{56}$$

To determine $p_Y(y)$, let $f(x)$ be quadratic in x, that is

$$f(x) = Ax^2 + Bx + C \tag{57}$$

By substituting for f in (56) the f of (57), and replacing I with the integral of f in (57), that is

$$I = A/3 + B/2 + C \tag{58}$$

one can obtain for $p_Y(y)$,

$$p_Y(y) = 6y(1-y) \tag{59}$$

It can be shown by substitution that the approximation I in (56), with $p_Y(y)$ of (59), is a zero variance estimator for the integral of quadratic integrands.

The function $f^*(x)$ could have been defined as

$$f^*(y) = f(1-y) - yf(0) - (1-y)f(1) \tag{60}$$

with a resulting estimator of

$$I_2 = \frac{f(0) + f(1)}{2} + \frac{f(1-y) - yf(0) - (1-y)f(1)}{6y(1-y)} \tag{61}$$

which is also exact for $f(x)$ quadratic in x. Since y and 1-y have the same distribution, but lie on the opposite side of 1/2, the average of (56) and (61), as can be shown by substitution, results in a zero variance estimator for $f(x)$ cubic in x[3]. Thus

$$I_3 = \frac{f(0) + f(1)}{2} + \frac{f(y) + f(1-y) - f(0) - f(1)}{12y(1-y)} \tag{62}$$

I_3 is exact for $f(x)$ cubic in x.

In the same manner as above, by defining f* as

$$f^*(x) = f(x)-(1-x)(1-2x)f(0)+x(1-2x)f(1)-4x(1-x)f(\tfrac{1}{2}) \tag{63}$$

a difference curve which is quadratic and passes through the two end points and the center of f(x) becomes the basis for the initial estimate for the integral. Applying the methods above, a zero variance estimator for quartic integrands becomes [3].

$$I_4 = \frac{f(0)+f(1)+4f(1/2)}{6} + \frac{f(y)-(1-y)(1-2y)f(0)+y(1-2y)f(1)-4y(1-y)f(1/2)}{30y(1-y)(1-2y)^2} \tag{64}$$

and a zero variance quintic estimator becomes 2

$$I_5 = \frac{f(0)+f(1)+4f(1/2)}{6} + \frac{f(y)+f(1-y)-(1-2y)^2[f(0)+f(1)]-8y(1-y)f(1/2)}{30y(1-y)(1-2y)^2} \tag{65}$$

The probability density function for both the quadratic and cubic zero variance estimators is (59) while that for the quartic and quintic zero variance estimators is

$$p_Y(y) = 30y(1-y)(1-2y)^2 \tag{66}$$

D. EXTENSIONS TO MULTIPLE INTEGRALS

The zero variance estimators above were derived in [2] only for single integrals. While the method can be extended to multiple integrals by nesting, nothing is gained since the

number of function evaluations increases exponentially with the number of variables of integration. In this section, double integral extensions to the quadratic, cubic, quartic, and quintic zero-variance estimators, and triple integral extensions to the quadratic and cubic estimators are proposed, and the method of derivation is presented. Although not thoroughly tested on a wide range of integrands, it is believed that this method of extension can, for multiple integrals with many variables of integration, provide an accurate means of approximating the integral with fewer evaluations of the integrand than for the quadrature sum routines.

For the double integral $\int_0^1 \int_0^1 f(x,y)\,dxdy$, consider the difference function which is a surface passing through the four corners $f(0, 0)$, $f(0, 1)$, $f(1, 0)$, and $f(1, 1)$. This function is

$$f(x,y)-f^*(x,y) = (1-x)(1-y)f(0,0)+(1-x)yf(0,1) + x(1-y)f(1,0) + xyf(1,1) \tag{67}$$

Integrating this difference function results in

$$\int_0^1 \int_0^1 [f(x,y)-f^*(x,y)]\,dxdy = \frac{f(0,0)+f(1.0)+f(0,1)+f(1,1)}{4} \tag{68}$$

Now if I^* is defined similar to (55) as

$$I^* = \frac{f^*(x,y)}{p_X(x)\,p_Y(y)} \tag{69}$$

and $p_X(x)$ and $p_Y(y)$ have the form of (59), solving for

$f^*(x,y)$ in (67) and using (69) as the estimate for $\int_0^1 \int_0^1 f^*(x,y)\,dxdy$, the estimate to $\int_0^1 \int_0^1 f(x,y)\,dxdy$ becomes

$$\int_0^1 \int_0^1 f(x,y)\,dxdy = \frac{f(0,0)+f(0,1)+f(1,0)+f(1,1)}{4}$$

$$+ \frac{f(x,y)-(1-x)(1-y)f(0,0)-(1-x)yf(0,1)-x(1-y)f(1,0)-xyf(1,1)}{36xy(1-x)(1-y)} \tag{70}$$

which is the double integral extension to (56). Replacing x with 1-x and y with 1-y in (70) and averaging the resulting expression with (70) gives the double integral extension to the cubic zero variance estimator (62)

$$\int_0^1 \int_0^1 f(x,y)\,dxdy = \frac{f(0,0)+f(0,1)+f(1,0)+f(1,1)}{4}$$
$$+ \frac{f(x,y)+f(1-x,1-y)-(1-x-y+2xy)[f(0,0)+f(1,1)]}{72xy(1-x)(1-y)}$$
$$+ \frac{-(x+y-2xy)[f(0,1)+f(1,0)]}{72xy(1-x)(1-y)} \tag{71}$$

For higher order extensions, a difference surface is defined such that it passes through the 2^k end points of $f(\underline{x}^T)$, where $\underline{x}^T$ is a k vector. This difference curve, which is a polynomial in each of its variables, can be integrated analytically to give a first approximation to the desired integral. The estimate is then refined by sampling the f* function weighted by the distribution of the variables, as in (69). Thus the triple extension to the quadratic estimator (56) becomes

$$\int_0^1 \int_0^1 \int_0^1 f(x,y,z)\,dxdydz = \frac{f(0,0,0)+f(0,0,1)+f(0,1,0)+f(0,1,1)}{8}$$

$$+ \frac{f(1,0,0)+f(1,0,1)+f(1,1,0)+f(1,1,1)}{8} + \frac{f(x,y,z)}{216xyz(1-x)(1-y)(1-z)}$$

$$+ \frac{-(1-x)(1-y)(1-z)f(0,0,0)-(1-x)(1-y)zf(0,0,1)}{216xyz(1-x)(1-y)(1-z)}$$

$$+ \frac{-(1-x)y(1-z)f(0,1,0)}{216xyz(1-x)(1-y)(1-z)}$$

$$+ \frac{-(1-x)yzf(0,1,1)-x(1-y)(1-z)f(1,0,0)-x(1-y)zf(1,0,1)}{216xyz(1-x)(1-y)(1-z)}$$

$$+ \frac{-xy(1-z)f(1,1,0)-xyzf(1,1,1)}{216xyz(1-x)(1-y)(1-z)} \qquad (72)$$

and the triple extension to the cubic estimator (62) becomes

$$\int_0^1 \int_0^1 \int_0^1 f(x,y,z)\,dxdydz = \frac{f(0,0,0)+f(0,0,1)+f(0,1,0)+f(0,1,1)}{8}$$

$$+ \frac{f(1,0,0)+f(1,0,1)+f(1,1,0)+f(1,1,1)}{8} + \frac{f(x,y,z)+f(1-x,1-y,1-z)}{432xyz(1-x)(1-y)(1-z)}$$

$$+ \frac{-(1-x-y-z+xz+yz)[f(0,0,0)+f(1,1,1)]}{432xyz(1-x)(1-y)(1-z)}$$

$$+ \frac{-(z-xz-yz+xy)[f(0,0,1)+f(1,1,0)]}{432xyz(1-x)(1-y)(1-z)}$$

$$+ \frac{-(y-xy-zy+xz)[f(0,1,0)+f(1,0,1)]}{432xyz(1-x)(1-y)(1-z)}$$

$$+ \frac{-(x-xz-xy+yz)[f(1,0,0)+f(0,1,1)]}{432xyz(1-x)(1-y)(1-z)} \qquad (73)$$

For the quartic and quintic double integral extensions, the difference surface must additionally pass through the four points which are peripheral midpoints and the internal midpoint of the integrand. The double integral extension for the quartic zero variance estimator becomes

$$\int_0^1 \int_0^1 f(x,y)\,dxdy = \frac{f(0,0)+f(0,1)+f(1,0)+f(1,1)}{36}$$

$$+ 4\left[\frac{f(0,\frac{1}{2})+f(1,\frac{1}{2})+f(\frac{1}{2},0)+f(\frac{1}{2},1)+4f(\frac{1}{2},\frac{1}{2})}{36}\right]$$

$$+ \frac{f(x,y)-(1-x)(1-2x)[(1-y)(1-2y)f(0,0)}{900xy(1-x)(1-y)(1-2x)^2(1-2y)^2}$$

$$\frac{+4y(1-y)f(0,\frac{1}{2})-y(1-2y)f(0,1)]}{900xy(1-x)(1-y)(1-2x)^2(1-2y)^2}$$

$$+ \frac{-4x(1-x)[(1-y)(1-2y)f(\frac{1}{2},0)+4y(1-y)f(\frac{1}{2},\frac{1}{2})-y(1-2y)f\frac{1}{2},1)]}{900xy(1-x)(1-y)(1-2x)^2(1-2y)^2}$$

$$+ \frac{x(1-2x)[(1-y)(1-2y)f(1,0)+4y(1-y)f(1,\frac{1}{2})-y(1-2y)f(1,1)]}{900xy(1-x)(1-y)(1-2x)^2(1-2y)^2} \qquad (74)$$

and the double integral extension for the quintic zero variance estimator becomes

$$\int_0^1 \int_0^1 f(x,y)\,dxdy = \frac{f(0,0)+f(0,1)+f(1,0)+f(1,1)}{36}$$

$$+ 4\left[\frac{f(0,\frac{1}{2})+f(1,\frac{1}{2})+f(\frac{1}{2},0)+f(\frac{1}{2},1)+4f(\frac{1}{2},\frac{1}{2})}{36}\right]$$

$$+ \frac{f(x,y)+f(1-x,1-y)-(1-2x)(1-2y)(1-x-y+2xy)[f(0,0)+f(1,1)]}{1800xy(1-x)(1-y)(1-2x)^2(1-2y)^2}$$

$$+ \frac{(1-2x)(1-2y)(x+y-2xy)[f(1,0)+f(0,1)]}{1800xy(1-x)(1-y)(1-2x)^2(1-2y)^2}$$

$$+ \frac{-4y(1-2x)(1-y)[f(0,\frac{1}{2})-f(1,\frac{1}{2})]}{1800xy(1-x)(1-y)(1-2x)^2(1-2y)^2}$$

$$+ \frac{-4x(1-x)(1-2y)[f(\frac{1}{2},0)-f(\frac{1}{2},1)]-32xy(1-x)(1-y)f(\frac{1}{2},\frac{1}{2})}{1800xy(1-x)(1-y)(1-2x)^2(1-2y)^2} \qquad (75)$$

For higher order extensions to the quartic and quintic estimators, the difference function must be defined so as to pass through the 2^k corners, the $k2^{k-1}$ peripheral midpoints, and the central midpoint of $f(\underline{x}^T)$, where $\underline{x}^T$ is a k vector. These extensions become difficult to derive due to the large number of points involved.

The triple integral extension to the cubic estimator, (73) has been applied to the example of Chapter II. The results are contained in Chapter IV for comparison, and a printout of the program is contained in Appendix B.

IV. CONCLUSIONS

Chapter II presents the mathematics for obtaining a performance measure's probability density function for explicit and implicit performance measures. Two integration schemes using Gauss quadrature sums were then discussed. Chapter III illustrated a crude Monte Carlo method for obtaining a performance measure's distribution and then discussed the confidence of such schemes. Monte Carlo schemes for evaluating single integrals were discussed. Double and triple integral extensions to these methods were then derived.

From observation of the problem, it seems intuitive that most system's performance measures will be of the implicit type, and in general the performance measure will be a function of several variables. In the process of evaluating the performance measure's distribution, the integral (1) must be evaluated. Evaluation by the quadrature sum methods, (25) and (29) require some n^k evaluations of the integrand, where n is the number of points in the quadrature sum, and k is the number of variables. This number can become quite large, and since the computation time for such routines is roughly proportional to the number of times the integrand is evaluated, this process could take a good deal of time. However, this method is quite accurate, and it is easily programmed for a large number of variables.

The reduced variance estimators reduce the number of evaluations of the integrand required. For a single estimate, the quintic extension requires evaluations at the 2^k corners, the $k2^{k-1}$ peripheral midpoints and the 3 internal points, or a total of $3 + (k+2)2^{k-1}$ points, where again k is the number of variables. For additional estimates, only two additional points per estimate are required. Thus M estimates of the integrand would require $2M + 1 + (k+2)2^{k-1}$ evaluations of the integrand. This can, in most cases be much less than that required for the quadrature sums.

The drawback to the reduced variance scheme is that the geometric form of the integrand must be examined to insure that the major portion of the surface will not be neglected in the sampling process. As illustrated in Figure 3, the difference curve does not provide a good representation of the function, while splitting the curve into two sections as in Figure 4 enables one to obtain a good representation of the function by the use of two difference curves.

The generation of the random numbers with the probability density function $30x(1-x)(1-2x)^2$ can be bothersome to program as can the quintic estimator extension. The cubic estimator is much more easy to extend, and the density function $6x(1-x)$ is more readily programmable. However more points are required by the cubic to ensure as good accuracy as the quintic extension.

Properly used, the Monte Carlo method of integral estimation can provide a more rapid method of determining a

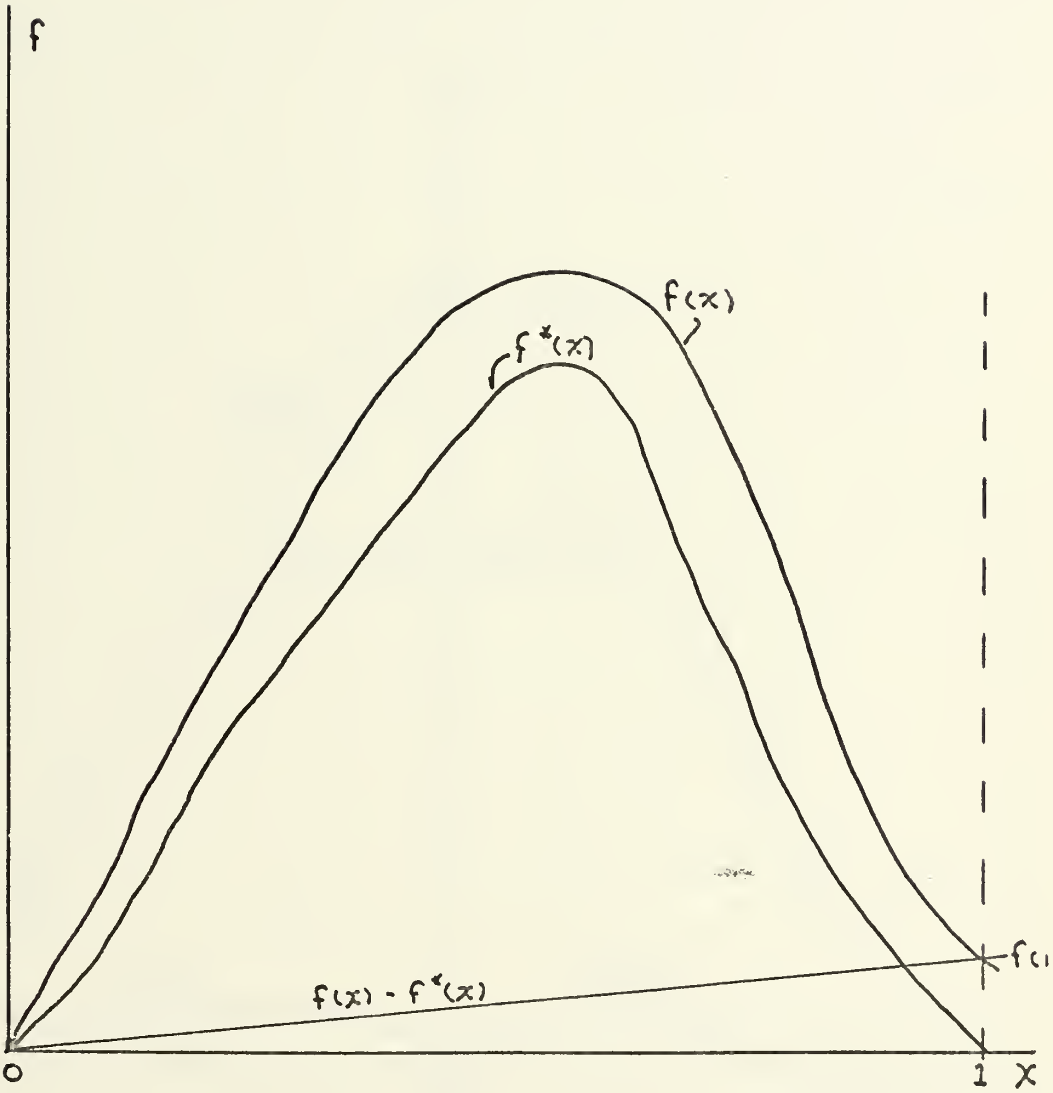

Figure 3. Improper Difference Curve Choice.

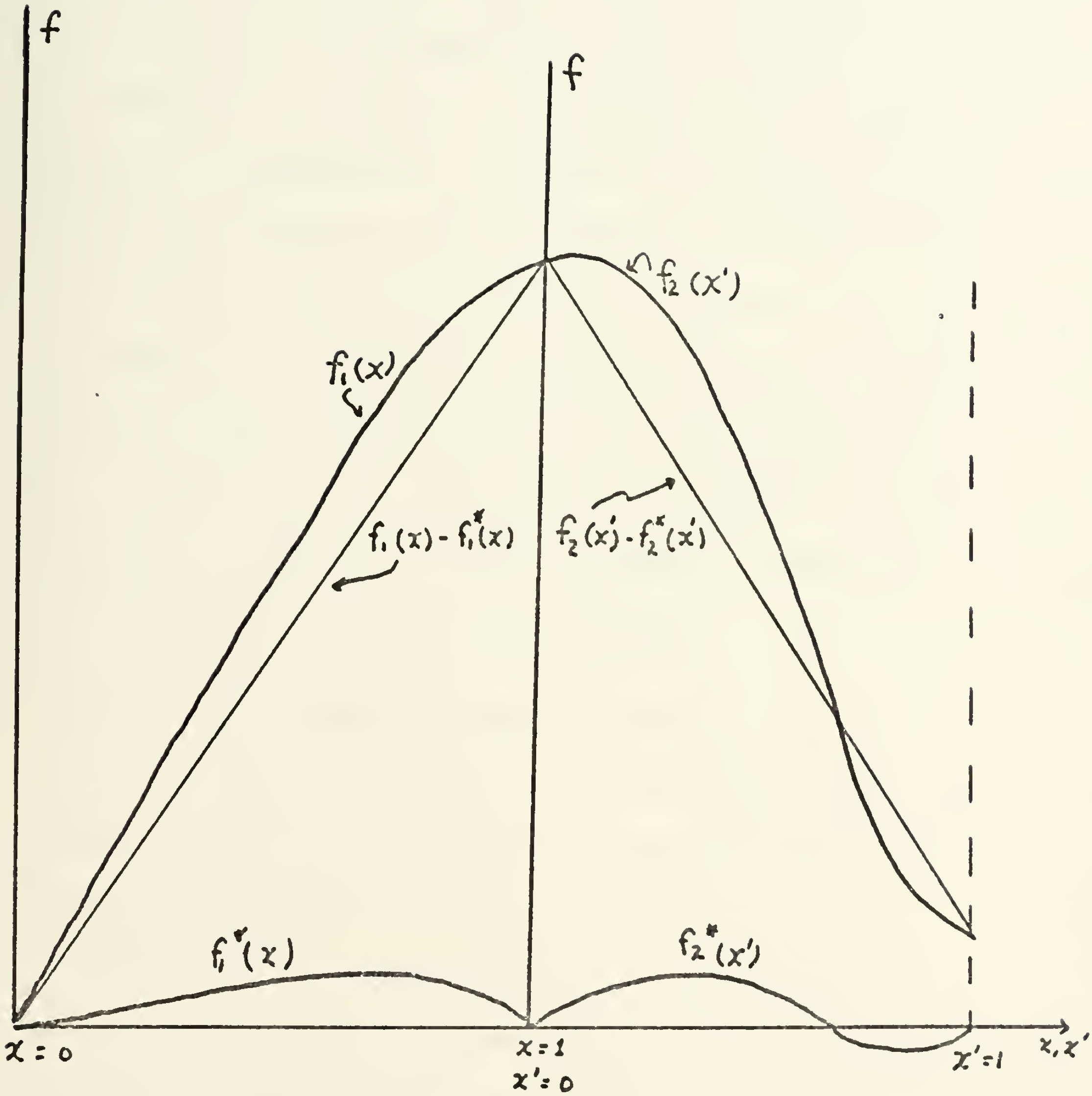

Figure 4. Proper Difference Curve Choice.

performance measure's density function with reasonable accuracy than other methods which require more standard integration routines.

A. RECOMMENDATIONS FOR FURTHER STUDY

Monte Carlo schemes seem to be attractive for reducing the amount of time required in the evaluation of the multiple integrals required in evaluation of a performance measure's distribution. Higher order zero variance estimators could be generated by fitting a difference curve to the points which are the zero's of a Legendre polynomial and evaluating the required probability density function to achieve a zero variance estimator for integrands of order 2n-1, where n is the order of the Legendre polynomial. Other schemes of generating random numbers of particular distributions could as well be investigated.

B. RESULTS OF APPLICATIONS TO EXAMPLE

The cumulative distribution function of the upper 3 db frequency of the series R-L-C circuit of Figure 1 has been obtained by four methods.

First, using the analytical expressions (11) and (12) for the integral (6), the probability density function, equation (15) was obtained by evaluating the integral with a 6-point Gauss Hermite quadrature sum routine. The cumulative distribution function was then obtained from the probability density function by rectangular integration. The resulting distribution has been plotted in Figure 5, and will serve as

the reference since integration of explicit functions is generally more accurate than integration of implicit functions.

Using the same performance measure as for the explicit case, but stipulating that the value of the performance measure was to be obtained implicitly, integral (18) was evaluated using 6-point Gauss Legendre quadrature sums. As can be seen in the plot of the cumulative distribution function for this case, Figure 6, the distribution is very nearly the same as that obtained in the explicit case.

Then for the crude Monte Carlo case, 1000 samples of the performance measure (10) were taken with the values of the three variables being obtained from a Gaussian random number generator. As can be seen in Figure 7, the resulting cumulative distribution function is very nearly the same as that obtained in the explicit case.

The extended cubic zero variance estimator equation (73) was then applied to the integral (18). The required distributions of the components were generated as explained in Appendix A. Twenty approximations for each value of the performance measure averaged to obtain the cumulative distribution plotted in Figure 8. Two discrepancies are observed here. First, the cumulative distribution exceeds the maximum value of 1, and secondly, there is an apparent discontinuity in the curve. Since the entire curve has a higher value than that of Figure 5, it is assumed that the difference surface estimation was too large.

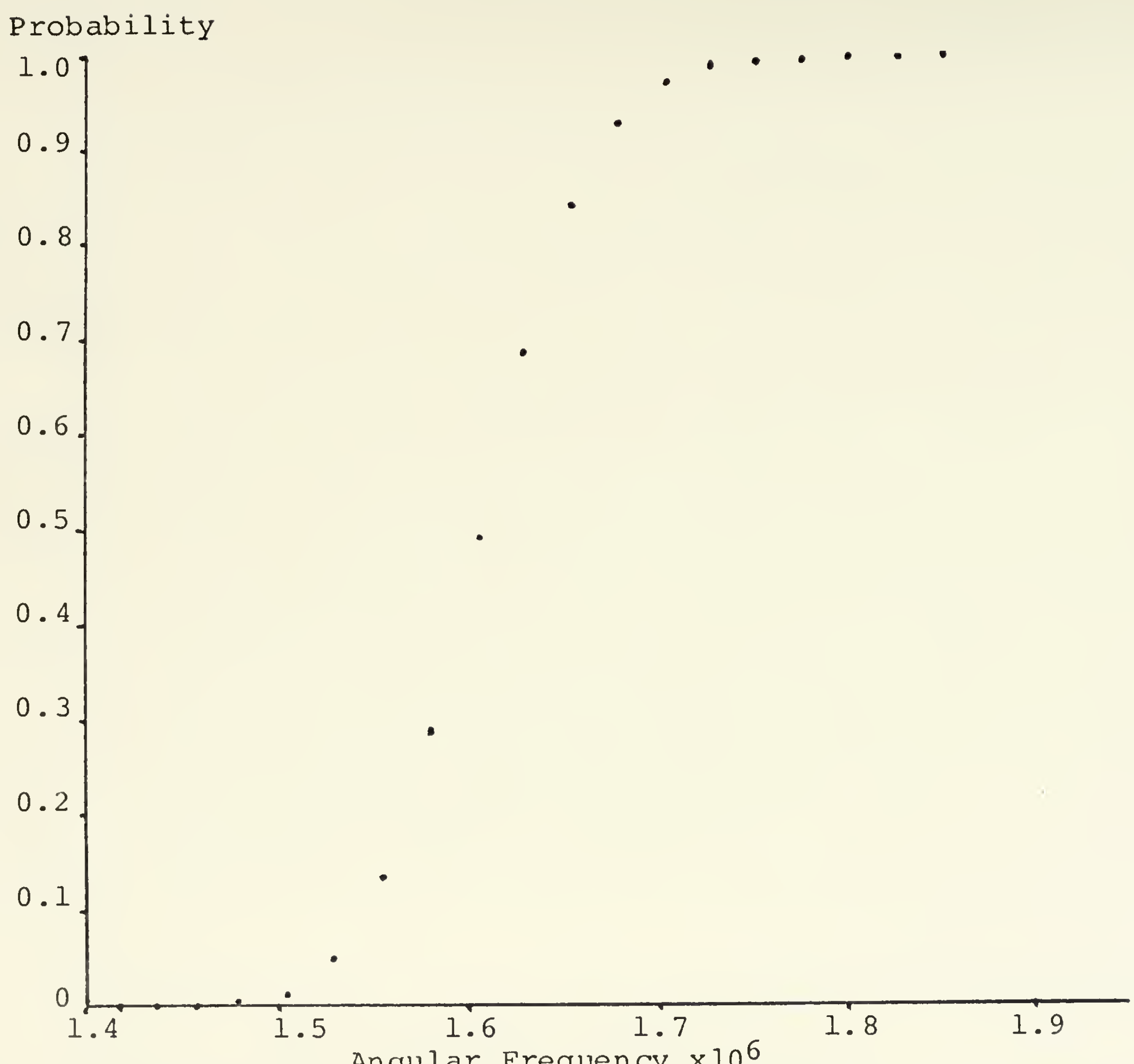

Figure 5. Cumulative Distribution Function for Example Using Gauss Hermite Quadrature (Explicit).

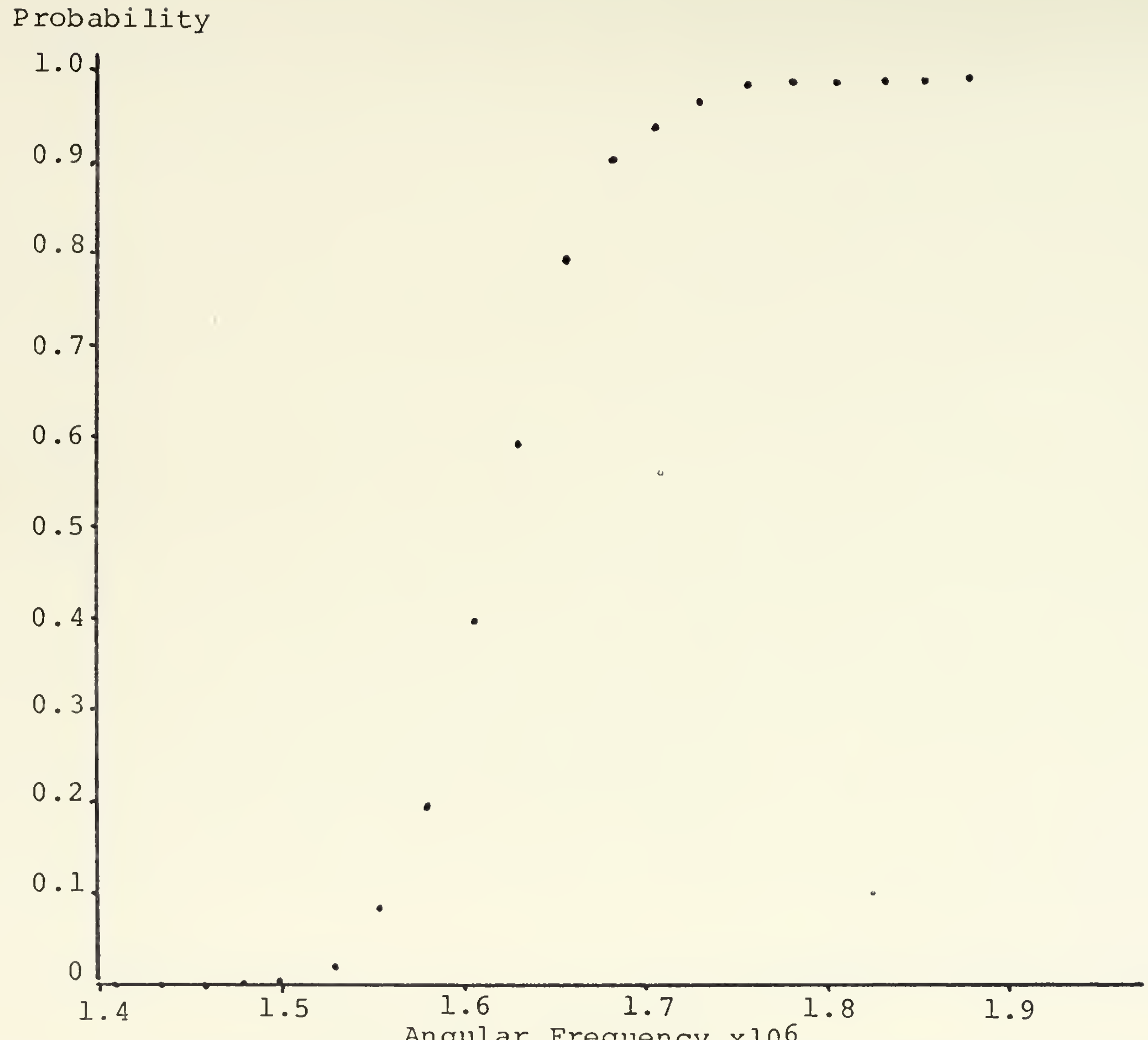

Figure 6. Cumulative Distribution Function for Example Using Gauss Legendre Quadrature (Implicit).

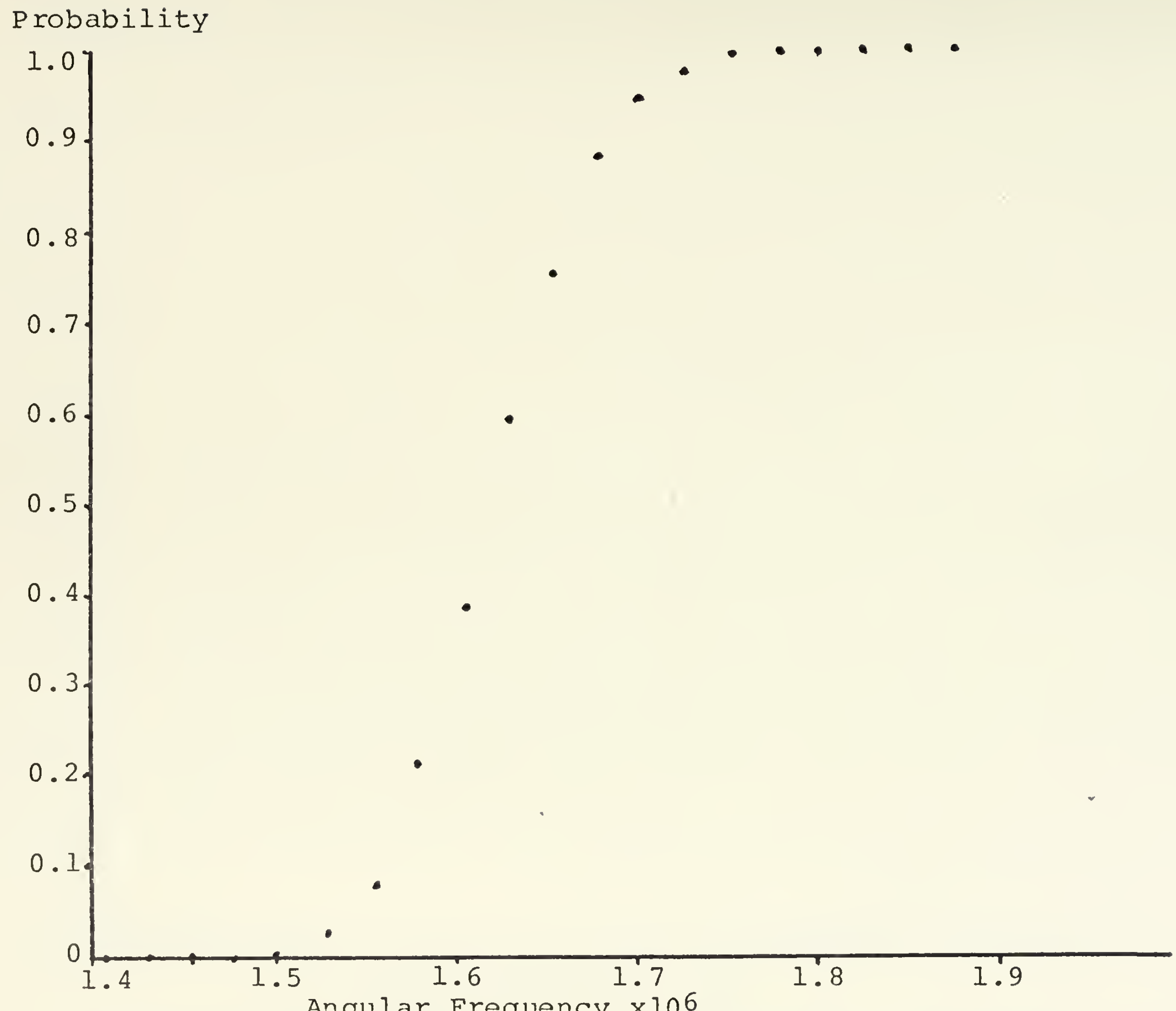

Figure 7. Cumulative Distribution Function for Example Using Crude Monte Carlo.

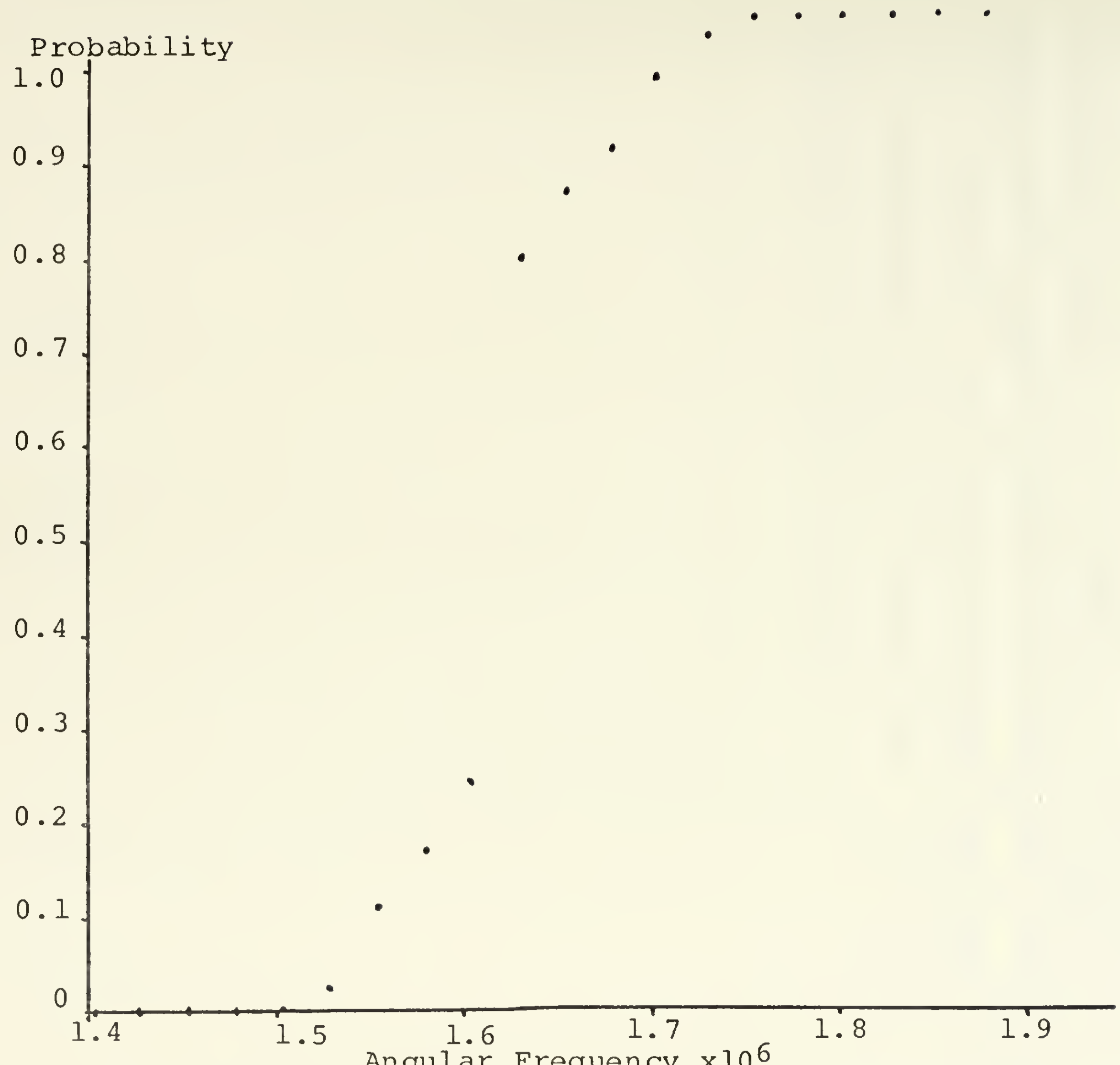

Figure 8. Cumulative Distribution Function for Example Using Extended Cubic Zero Variance Estimator.

Although the example did not prove the worth of the Monte Carlo method, a count of the required number of evaluations for high multiple integrands for a 6-point Gauss Legendre routine as compared to that required in the extended cubic estimator averaging 100 approximations for each value of the performance measure will illustrate its merits. For a multiple integral with 10 variables of integration, the 6-point Gauss Legendre method would require 6^{10} evaluations of the integrand. For the same integrand, the cubic extension would require $200 + 2^{10}$ evaluations. The savings in computation time should be apparent.

Appendix A

RANDOM NUMBER GENERATION

A. UNIFORM RANDOM NUMBERS

A truly uniform random number cannot be generated on a digital computer. However, methods such as the power residue [4, 10] method can generate a sequence of numbers which have the properties of uniformly distributed random numbers.

The method of power residues is based on modular arithmetic. If X_i is an arbitrary integer in the sequence of pseudo-random numbers to be generated, the next number generated is determined by

$$X_{i+1} = X_i p \text{ (modulo M)} \qquad \text{A-1)}$$

where p is any prime integer such that p^2 is greater than M, the modular base of the computer's arithmetic unit. The number X_i/M is then an approximation to a random variable with a uniform distribution on the interval (0,1). Utilizing a digital computer with a 32 bit register (31 number bits plus one sign bit), a value of p of 65,539 for an M of 2,147,483,647 = $2^{31}-1$, a sequence of 2^{29} numbers can be generated before any number of the sequence is repeated [4].

Since the digital computer automatically performs modular arithmetic on integers, with the modulus being determined by the size of the registers in the arithmetic unit, uniform pseudo random numbers can be generated quite rapidly.

B. GAUSSIAN RANDOM NUMBERS

Gaussian random numbers are most commonly generated utilizing uniformly distributed pseudo-random numbers. If R_i, $i=1,2,\ldots,n$ are independent samples from a density function uniformly distributed over the interval (0,1), then a normally distributed random variable with 0 mean and variance of 12/n can be approximated by [3, 4, 10]

$$Z = \frac{12}{n} \sum_{i=1}^{n} (R_i) - 6 \tag{A-2}$$

If n = 12, Z will have 0 mean and variance of 1.

C. 6x(1-x) DISTRIBUTION

If R_i is an independent sample from a random variable uniformly distributed on the interval (0,1), for i = 1,2,3, and if the three samples are rearranged in order of magnitude such that

$$R_1 \leq R_2 \leq R_3 \tag{A-3}$$

then R_2 has probability density function $6x(1-x)$ [3].

D. $30x(1-x)(1-2x)^2$ DISTRIBUTION

If R_1, R_2, R_3, R_4, and R_5 are independent samples from a random variable uniformly distributed over the interval (0,1), and are arranged so that

$$|R_1 - \tfrac{1}{2}| \leq |R_2 - \tfrac{1}{2}| \leq \ldots \leq |R_5 - \tfrac{1}{2}| \tag{A-4}$$

and if S is chosen such that $S = R_4$ with probability 3/4 and $S = R_3$ with probability 1/4, then S has probability density

function $30x(1-x)(1-2x)^2$ [3]. This density function, it was discovered, requires some complex programming to generate.

E. OTHER DISTRIBUTIONS

Let Y be a uniformly distributed random variable over the interval (0,1), and let

$$y = F(x) = \int_{-\infty}^{Y} h(t)\,dt \qquad \text{(A-5)}$$

where h(t) = 1 for 0 t 1, and h(t) = 0 otherwise. Since

$$F(x) = \int_{-\infty}^{x} p_X(x) \qquad \text{(A-6)}$$

where $p_X(x)$ is the probability density function for X, if F(x) or $p_X(x)$, the desired function, are known explicitly, then the inverse function

$$x = F^{-1}(y) \qquad \text{(A-7)}$$

will generate x according to its desired density function $p_X(x)$ [3]. Most often $p_X(x)$ will be a polynomial fitted to some statistical data which represents the distribution of the variable X. Since this polynomial is normally of a high order, some difficulty can result from this method in that the zeros of (A-6) may be difficult to find.

APPENDIX B

COMPUTER PROGRAMS

```
C     COMPUTATION OF CUMULATIVE DISTRIBUTION FUNCTION OF 3
C     DB ANGULAR FREQUENCY OF THE SERIES R-L-C CIRCUIT
C     EXAMPLE.  PERFORMANCE MEASURE IS AN EXPLICIT FUNCTION
C     OF THE VARIABLES.  METHOD IS 6 POINT GAUSS HERMITE
C     QUADRATURE SUM APPROXIMATION TO EQUATION (15) TO
C     OBTAIN PROBABILITY DENSITY FUNCTION.  RECTANGULAR
C     INTEGRATION IS THEN USED TO OBTAIN CUMULATIVE
C     DISTRIBUTION FUNCTION.
      COMMON Z
      DIMENSION A(6),Y(6),X(2)
      WRITE(6,61)
      F=0.
C     INPUT GAUSS HERMITE COEFFICIENTS AND ZEROS
      A(1)=.453001E-2
      A(2)=.1570607
      A(3)=.7246296
      A(4)=A(3)
      A(5)=A(2)
      A(6)=A(1)
      Y(1)=2.350605
      Y(2)=1.335849
      Y(3)=.4360741
      Y(4)=-Y(3)
      Y(5)=-Y(2)
      Y(6)=-Y(1)
C     INPUT WORST CASE FREQUENCIES AND COMPUTE DELTA Z AND
C     DISCRETE FREQUENCIES.
      ZHI=.187918557
      ZLO=.140598767
      DELZ=(ZHI-ZLO)/19.
      Z=ZLO
C     FOR EACH Z, EVALUATE PDF USING 6 POINT GAUSS HERMITE
C     QUADRATURE.
      DO 3 IZ=1,20
      P=0.
      DO 2 I=1,6
      DO 2 J=1,6
      X(1)=Y(J)
      X(2)=Y(I)
    2 P=P+A(I)*A(J)*FUN(X)
C     USE RECTANGULAR INTEGRATION NOW TO OBTAIN CUMULATIVE
C     DISTRIBUTION.
      F=F+P*DELZ
C      OUTPUT CUMULATIVE DISTRIBUTION
      WRITE(6,60)Z,F
C       INCREMENT Z
    3 Z=Z+DELZ
      STOP
   60 FORMAT(2F30.7)
   61 FORMAT('1',////////,15X,'ANGULAR FREQUENCY',16X,
     1'DISTRIBUTION',//)
      END
```

```
      FUNCTION FUN(X)
C     FUNCTION SUBROUTINE FOR USE WITH GAUSS HERMITE
C     QUADRATURE SUM ROUTINE.
      COMMON Z
      DIMENSION X(2),XO(3),SGMA(3),X1(3)
      FUN=0.
C     CHECK FLAG TO SEE IF CONSTANTS HAVE BEEN COMPUTED.
      IF(II-121)1,3,1
    1 PI=SQRT(3.14159)**3
      P2=SQRT(2.)
C     INPUT MEAN VALUES OF COMPONENTS
      XO(1)=1.E-3
      XO(2)=1.E-9
      XO(3)=1.E3
C     CALCULATE STANDARD DEVIATIONS OF COMPONENTS
      DO 2 I=1,3
    2 SGMA(I)=XO(I)/30.
C     SET FLAG
      II=121
    3 CONTINUE
      DO 4 I=1,2
C     "UNNORMALIZE" VARIABLES FOR COMPUTATION OF DERIVATIVE
C     AND FUNCTION INVERSE.
    4 X1(I)=X(I)*P2*SGMA(I)+XO(I)
      X1(3)=Z*X1(1)-1./(Z*X1(2))
      DRV=X1(1)+1./(Z*Z*X1(2))
C     COMPUTE EXPONENT OF EXPONENTIAL
      F=(X1(3)-XO(3))/(P2*SGMA(3))
      F=F*F
      IF(F-13.6)5,6,6
C     EVALUATE FUNCTION
    5 FUN=DRV*EXP(-F)/(P2*PI*SGMA(3))
    6 RETURN
      END
```

ANGULAR FREQUENCY	DISTRIBUTION
0.1405987E 07	0.9358540E-06
0.1430892E 07	0.2585363E-04
0.1455797E 07	0.3094762E-03
0.1480702E 07	0.2316327E-02
0.1505607E 07	0.1303275E-01
0.1530512E 07	0.4960799E-01
0.1555417E 07	0.1364721E 00
0.1580322E 07	0.2939899E 00
0.1605227E 07	0.4935315E 00
0.1630132E 07	0.6891426E 00
0.1655037E 07	0.8430275E 00
0.1679942E 07	0.9314869E 00
0.1704847E 07	0.9743903E 00
0.1729752E 07	0.9925191E 00
0.1754657E 07	0.9980668E 00
0.1779562E 07	0.9994828E 00
0.1804467E 07	0.9998813E 00
0.1829372E 07	0.9999751E 00
0.1854277E 07	0.9999889E 00
0.1879182E 07	0.9999900E 00

```
C     COMPUTATION OF CUMULATIVE DISTRIBUTION FUNCTION OF 3
C     DB ANGULAR FREQUENCY OF THE SERIES R-L-CLC CIRCUIT
C     EXAMPLE.  PERFORMANCE MEASURE IS IMPLICIT FUNCTION OF
C     VARIABLES.  METHOD IS 6 POINT GAUSS LEGENDRE
C     QUADRATURE SUM APPROXIMATION TO EQUATION (18).
      COMMON Z(20),IZ
      DIMENSION A(6),Y(6),X(3),F(20)
      WRITE(6,61)
C     INPUT GAUSS LEGENDRE COEFFICIENTS AND ZEROS.
      A(1)=.1713245
      A(2)=.3607616
      A(3)=.4679139
      A(4)=A(3)
      A(5)=A(2)
      A(6)=A(1)
      Y(1)=.9324695
      Y(2)=.6612094
      Y(3)=.2386192
      Y(4)=-Y(3)
      Y(5)=-Y(2)
      Y(6)=-Y(1)
C     INPUT WORST CASE FREQUENCIES AND COMPUTE DELTA Z AND
C     DISCRETE FREQUENCIES.
      ZHI=.18791857
      ZLO=.14059877
      DELZ=(ZHI-ZLO)/19.
      Z(1)=ZLO
      DO 5 IZ=1,20
      IF(IZ.EQ.1) GO TO 1
      Z(IZ)=Z(IZ-1)+DELZ
    1 F(IZ)=0.
C     FOR EACH FREQUENCY, COMPUTE 6 POINT GAUSS LEGENDRE SUM
      DO 2 I=1,6
      DO 2 J=1,6
      DO 2 K=1,6
      X(1)=Y(K)
      X(2)=Y(J)
      X(3)=Y(I)
      A1=A(I)*A(J)*A(K)
    2 F(IZ)=F(IZ)+A1*FUN(X)
C     OUTPUT DISTRIBUTION
    5 WRITE(6,60)Z(IZ),F(IZ)
      STOP
   60 FORMAT(2E30.7)
   61 FORMAT('1',////////,15X,'ANGULAR FREQUENCY',16X,
     1'DISTRIBUTION',//)
      END
```

```
      FUNCTION FUN(X)
C     FUNCTION SUBROUTINE FOR USE WITH GAUSS LEGENDRE
C     INTEGRATION SCHEME.
      COMMON Z(20),IZ
      DIMENSION XO(3),SGMA(3),Y(3)
      DIMENSION X(3)
      PI=SQRT(2.*3.14159)**3
C     INPUT MEAN VALUES OF COMPONENTS
      XO(1)=1.E3
      XO(2)=1.E-3
      XO(3)=1.E-9
C     CALCULATE COMPONENT DEVIATIONS
      DO 4 I=1,3
      SGMA(I)=XO(I)/30.
C     "UNNORMALIZE" VARIABLES
    4 Y(I)=3.*SGMA(I)*X(I)+XO(I)
      P=Y(1)/(2.*Y(2))
C     COMPUTE VALUE OF PERFORMANCE MEASURE FOR THESE
C     COMPONENT VALUES.
      ZZ=P+SQRT(P*P+1./(Y(2)*Y(3)))
C     CHECK TO SEE IF PERFORMANCE MEASURE IS LESS THAN
C     REFERENCE VALUE.
      IF(ZZ.LE.Z(IZ))GO TO 1
C     IF NOT, THE INTEGRAND IS ZERO
      FUN=0.
      GO TO 3
C     IF IT IS COMPUTE VALUE OF INTEGRAND
    1 F=0.
      DO 2 I=1,3
    2 F=F+X(I)**2
      FUN=27.*EXP(-4.5*F)/PI
    3 RETURN
      END
```

ANGULAR FREQUENCY	DISTRIBUTION
0.1405987E 07	0.0
0.1430892E 07	0.1082632E-05
0.1455797E 07	0.4025977E-04
0.1480702E 07	0.6114664E-03
0.1505607E 07	0.5107399E-02
0.1530512E 07	0.2215009E-01
0.1555417E 07	0.8334792E-01
0.1580322E 07	0.1960930E 00
0.1605227E 07	0.3981495E 00
0.1630132E 07	0.5892813E 00
0.1655037E 07	0.7912316E 00
0.1679942E 07	0.9019774E 00
0.1704847E 07	0.9383603E 00
0.1729752E 07	0.9685229E 00
0.1754657E 07	0.9841329E 00
0.1779562E 07	0.9867263E 00
0.1804467E 07	0.9874057E 00
0.1829372E 07	0.9874438E 00
0.1854277E 07	0.9874458E 00
0.1879182E 07	0.9874459E 00

```
C     COMPUTATION OF CUMULATIVE DISTRIBUTION FUNCTION OF 3
C     DB ANGULAR FREQUENCY OF THE SETIES R-L-C CIRCUIT
C     EXAMPLE.  METHOD IS CRUDE MONTE CARLO, WHERE THE
C     PERFORMANCE MEASURE IS SAMPLED 1000 TIMES.  COMPONENT
C     VALUES ARE OBTAINED FROM GAUSSIAN RANDOM NUMBER
C     GENERATORS.
      DIMENSION IX(3)
      DIMENSION Z(20),X(3),XO(3),SGMA(3),F(20)
      WRITE(6,61)
C     INPUT WORST CASE FREQUENCIES AND COMPUTE DELTA Z
      ZLO=.14C598E7
      ZHI=.1879185E7
      DELZ=(ZHI-ZLO)/19.
      Z(1)=ZLO
C     INITIALIZE DISTRIBUTION TO ZERO.
      F(1)=0.
      DO 1 I=2,20
      Z(I)=Z(I-1)+DELZ
    1 F(I)=0.
C     INPUT MEAN VALUES OF VARIABLES AND COMPUTE DEVIATION.
      XO(1)=1.E3
      XO(2)=1.E-3
      XC(3)=1.E-9
      DO 2 I=1,3
    2 SGMA(I)=XO(I)/30.
C     INPUT KERNALS FOR RANDOM GENERATION
      IX(1)=13495
      IX(2)=54621
      IX(3)=174327
C     SAMPLE PERFORMANCE MEASURE 1000 TIMES
      DO 4 I=1,1000
C     GENERATE COMPONENT VALUES FROM GAUSSIAN DISTRIBUTIONS.
      DO 3 J=1,3
    3 CALL GRN(IX(J),SGMA(J),XO(J),X(J))
C     COMPUTE PERFORMANCE MEASURE.
      PP=X(1)/(2.*X(2))
      ZZ=PP+SQRT(PP**2+1./(X(2)*X(3)))
C     FIND DISCRETE VALUES OF Z FOR WHICH PERFORMANCE
C     MEASURE IS LESS THAN OR EQUAL TO, AND INCREMENT THE
C     DISTRIBUTION COUNTERS.
      DO 4 J=1,20
      IF(ZZ.LE.Z(J))F(J)=F(J)+1.
    4 CONTINUE
C     DIVIDE DISTRIBUTION COUNTER BY 1000 TO OBTAIN
C     DISTRIBUTION
      DO 6 J=1,20
      F(J)=F(J)*1.E-3
C     OUTPUT DISTRIBUTION.
    6 WRITE(6,60)Z(J),F(J)
      STOP
   60 FORMAT(2E30.7)
   61 FORMAT('1',////////,15X,'ANGULAR FREQUENCY',16X,
     1'DISTRIBUTION',//)
      END
```

```
      SUBROUTINE GRN(IX,S,AM,V)
C     SUBROUTINE TO GENERATE GAUSSIAN DISTRIBUTED RANDOM
C     VARIABLES.
      A=0.
      DO 1 I=1,12
      CALL URN(IX,IY,Y)
      IX=IY
    1 A=A+Y
      V=(A-6.)*S+AM
      RETURN
      END

      SUBROUTINE URN(IX,IY,YFL)
C     SUBROUTINE TO GENERATE UNIFORM DISTRIBUTED RANDOM
C     VARIABLES.
      IY=IX*65539
      IF(IY)1,2,2
    1 IY=IY+2147483647+1
    2 YFL=IY
      YFL=YFL*.4656613E-9
      RETURN
      END
```

ANGULAR FREQUENCY	DISTRIBUTION
0.1405980E 07	0.0
0.1430885E 07	0.0
0.1455790E 07	0.0
0.1480695E 07	0.0
0.1505600E 07	0.6999999E-02
0.1530505E 07	0.2600000E-01
0.1555410E 07	0.7799995E-01
0.1580315E 07	0.2120000E 00
0.1605220E 07	0.3849999E 00
0.1630125E 07	0.5929999E 00
0.1655030E 07	0.7549999E 00
0.1679935E 07	0.8839999E 00
0.1704840E 07	0.9489999E 00
0.1729745E 07	0.9789999E 00
0.1754650E 07	0.9949999E 00
0.1779555E 07	0.9999999E 00
0.1804460E 07	0.9999999E 00
0.1829365E 07	0.9999999E 00
0.1854270E 07	0.9999999E 00
0.1879175E 07	0.9999999E 00

```
C     COMPUTATION OF CUMULATIVE DISTRIBUTION FUNCTION OF 3
C     DB ANGULAR FREQUENCY OF SERIES R-L-C CIRCUIT EXAMPLE.
C     PERFORMANCE MEASURE IS AN IMPLICIT FUNCTION OF ITS
C     VARIABLES.  METHOD IS TRIPLE INTEGRAL EXTENSION TO
C     CUBIC ZERO VARIANCE ESTIMATOR APPLIED TO EQUATION (18)
      COMMON Z(20),IZ
      DIMENSION X(3),X1(3),Y(3),IX(3),IX1(3),F(20)
      WRITE(6,61)
C     INPUT WORST CASE FREQUENCIES AND COMPUTE DELTA Z AND
C     DISCRETE VALUES OF Z.
      ZHI=.18791857
      ZLO=.14059877
      DELZ=(ZHI-ZLO)/19.
      Z(1)=ZLO
      DO 1 I=2,20
    1 Z(I)=Z(I-1)+DELZ
C     DO FOLLOWING FOR EACH REFERENCE VALUE OF Z
      DO 999 IZ=1,20
C     INPUT KERNALS FOR RANDOM NUMBER GENERATION.
      IX(1)=49431
      IX(2)=394163
      IX(3)=1755381
      SUM=0.
C     COMPUTE "CORNER VALUES" OF INTEGRAND AND GROUP INTO
C     FOUR VALUES.
      DO 2 I=1,3
      X(I)=0.
    2 X1(I)=1.-X(I)
      F1=FUN(X)+FUN(X1)
      X(3)=1.
      X1(3)=0.
      F2=FUN(X)+FUN(X1)
      X(1)=1.
      X1(1)=0.
      F3=FUN(X)+FUN(X1)
      X(3)=0.
      X1(3)=1.
      F4=FUN(X)+FUN(X1)
C     COMPUTE INITIAL ESTIMATE
      AVG=(F1+F2+F3+F4)/8.
C     NOW GET THE AVERAGE OF 20 REFINEMENTS
      DO 7 JX=1,20
C     GENERATE 3 RANDOM VARIABLES WITH 6X(1-X) DISTRIBUTION.
      DO 800 J=1,3
      DO 6 I=1,3
      CALL URN(IX(J),IX1(J),Y(I))
    6 IX(J)=IX1(J)
      IF(Y(1)-Y(2))100,1000,200
  100 IF(Y(1)-Y(3))700,1000,500
  200 IF(Y(2)-Y(3))300,1000,400
  300 IF(Y(1)-Y(3))500,1000,600
  400 X(J)=Y(2)
      GO TO 800
  500 X(J)=Y(1)
      GO TO 800
  600 X(J)=Y(3)
      GO TO 800
  700 IF(Y(2)-Y(3))400,1000,600
  800 X1(J)=1.-X(J)
C     COMPUTE REFINEMENT
      SUM1=0.
      SUM1=SUM1+FUN(X)+FUN(X1)
      SUM1=SUM1-F1*(1.-X(1)-X(2)-X(3)+X(1)*X(2)+X(1)*X(3)+
     1X(2)*X(3))
      SUM1=SUM1-F2*(X(3)-X(1)*X(3)-X(2)*X(3)+X(1)*X(2))
      SUM1=SUM1-F3*(X(2)-X(1)*X(2)-X(3)*X(2)+X(1)*X(3))
      SUM1=SUM1-F4*(X(1)-X(1)*X(2)-X(1)*X(3)+X(2)*X(3))
    7 SUM=SUM+SUM1/(432.*X(1)*X(2)*X(3)*X1(1)*X1(2)*X1(3))
C     COMPUTE THE VALUE OF CUMULATIVE DISTRIBUTION FOR THIS
C     VALUE OF Z
      F(IZ)=.05*SUM+AVG
      F(IZ)=F(IZ)/1.125
```

```
C     OUTPUT DISTRIBUTION
  999 WRITE(6,60)Z(IZ),F(IZ)
 1000 STOP
   60 FORMAT(2E30.7)
   61 FORMAT('1',////////,15X,'ANGULAR FREQUENCY',16X,
     1'DISTRIBUTION',//)
      END

      FUNCTION FUN(X)
C     FUNCTION SUBROUTINE FOR MONTE CARLO METHOD
      COMMON Z(20),IZ
      DIMENSION XO(3),SGMA(3),Y(3),X(3)
C     CHECK FLAG TO DETERMINE IF CONSTANTS HAVE BEEN
C     EVALUATED.
      IF(II-114)1,3,1
    1 PI=SQRT(2.*3.14159)**3
C     INPUT MEAN VALUES OF COMPONENTS
      XO(1)=1.E3
      XO(2)=1.E-3
      XO(3)=1.E-9
C     COMPUTE DEVIATIONS OF COMPONENTS.
      DO 2 I=1,3
    2 SGMA(I)=XO(I)/30.
C     SET FLAG
      II=114
    3 G=0.
C     BREAK INTEGRAND INTO 8 PARTS AND SUM INDIVIDUAL
C     EVALUATIONS.
      DO 6 I=1,2
C     "UNNORMALIZE" VARIABLES FOR EACH PART.
      Y(1)=(-1)**I*3.*SGMA(1)*X(1)+XO(1)
      DO 6 J=1,2
      Y(2)=(-1)**J*3.*SGMA(2)*X(2)+XO(2)
      DO 6 K=1,2
      Y(3)=(-1)**K*3.*SGMA(3)*X(3)+XO(3)
      P=Y(1)/(2.*Y(2))
C     COMPUTE VALUE OF PERFORMANCE MEASURE
      ZZ=P+SQRT(P*P+1./(Y(2)*Y(3)))
C     CHECK VALUE OF PERFORMANCE MEASURE AGAINST REFERENCE.
      IF(ZZ-Z(IZ))4,4,6
C     IF LESS THAN OR EQUAL TO REFERENCE, EVALUATE FUNCTION
    4 F=0.
      DO 5 L=1,3
    5 F=F+X(L)**2
      G=G+EXP(-4.5*F)
C     IF GREATER THAN REFERENCE, SET FUNCTION EQUAL TO ZERO.
    6 CONTINUE
      FUN=27.*G/PI
      RETURN
      END

      SUBROUTINE URN(IX,IY,YFL)
C     SUBROUTINE FOR GENERATING UNIFORM RANDOM NUMBERS.
      IY=IX*65539
      IF(IY)1,2,2
    1 IY=IY+2147483647+1
    2 YFL=IY
      YFL=YFL*.4656613E-9
      RETURN
      END
```

ANGULAR FREQUENCY	DISTRIBUTION
0.1405987E 07	0.8984364E-07
0.1430892E 07	0.8984364E-07
0.1455797E 07	0.3900373E-04
0.1480702E 07	0.4555574E-03
0.1505607E 07	0.2644415E-02
0.1530512E 07	0.2990580E-01
0.1555417E 07	0.1155540E 00
0.1580322E 07	0.1728926E 00
0.1605227E 07	0.2506810E 00
0.1630132E 07	0.8067204E 00
0.1655037E 07	0.8746661E 00
0.1679942E 07	0.9280446E 00
0.1704847E 07	0.1000616E 01
0.1729752E 07	0.1047179E 01
0.1754657E 07	0.1061230E 01
0.1779562E 07	0.1062497E 01
0.1804467E 07	0.1062828E 01
0.1829372E 07	0.1062848E 01
0.1854277E 07	0.1062867E 01
0.1879182E 07	0.1062867E 01

LIST OF REFERENCES

1. Davis, P. J. and Rabinowitz, P., Numerical Integration, Blaisdell, 1967.

2. Franson, A. L., Prediction of Statistical System Performance from Parameter Distributions, M.S. Thesis, Naval Postgraduate School, Monterey, California, 1969.

3. Hammersley, J. M. and Handscomb, D. C., Monte Carlo Methods, Metheun and Co., 1964.

4. I.B.M., Generation of Random Numbers, I.B.M. Reference Manual C-20-8011, 1963.

5. Krylov, V. I., Approximate Calculation of Integrals, Macmillan, 1962.

6. Mark, D. G. and Stember, Jr., L. H., "Variability Analysis," Electro-Technology, p. 37-48, July 1965.

7. Meyer, P. L., Introductory Probability and Statistical Applications, Addison-Wesley, 1965.

8. Milne, E. M., Numerical Calculus, Princeton University Press, 1949.

9. Scarborough, J. B., Numerical Mathematical Analysis, Johns Hopkins Press, 1966.

10. Shrieder, Y. A., The Monte Carlo Method, Pergammon Press, 1966.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Professor S. R. Parker, Code 52 Pz Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	1
4.	LT Dennis R. Oldson c/o Commander Service Forces Pacific FPO, San Francisco, California 96610	1

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

STATISTICAL EXTRAPOLATION

4. DESCRIPTIVE NOTES *(Type of report and, inclusive dates)*
Engineer's Degree Thesis; September 1971.

5. AUTHOR(S) *(First name, middle initial, last name)*

Dennis Randolph Oldson
Lieutenant, United States Navy

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1971	67	10

8a. CONTRACT OR GRANT NO. b. PROJECT NO c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

The mathematics of extrapolating known statistics of components to the probability density function of a system's performance measure is considered. Quadrature sum integration schemes for evaluating the resulting required integration are examined, and alternate integral approximation schemes are developed utilizing Monte Carlo methods. A simple electrical circuit example illustrates the use of these techniques.

Unclassified
Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Performance Measure						
Extrapolation						
Monte Carlo Methods						
Probability Distribution Function						

DD FORM 1 NOV 65 1473 (BACK) Unclassified
0101-807-6821 Security Classification A-31409

132919
Thesis
0382 Oldson
c.1 Statistical extra-
polation.